University of Essex

Predicting Life Expectancy

- Group 50 -

- MA317: Modelling Experimental Data -

- 27.03.2022 -

Summary

The project consists of a short analysis of the World Development Indicators (WDI) dataset to predict life expectancy at birth (in years) for the year 2019. Moreover, it aims to study if there is any significant difference in the average life expectancy across multiple continents. The main findings suggest that the **infant mortality rate**, **health expenditure per capita** and **the percentage of people using safely managed drinking water services** are good predictors for life expectancy at birth.

Contents

Remark: The titles above contain hyperlinks to the relevant sections, which can be clicked.

Word count for the main analysis: 2707

Introduction

Life Expectancy at birth is a good indicator for the overall development level of a country and therefore, it is an important statistic to keep track of. However, it is similarly important to identify the driving factors behind it.

This project aims to propose the best model for predicting life expectancy at birth. The main assumption that the analysis will be based on is that there exists a linear relationship between the predictor variables and the predicted variable (*Life expectancy at birth*).

Naturally, the project will conclude with a linear regression model that will be used to estimate the value of life expectancy for the countries where this data is missing.

Remark: To manipulate the columns more easily, their names were changed and the meaning of each of them can be found in the appendix [1]. The variable Life Expectancy at birth will be referred to as **y**.

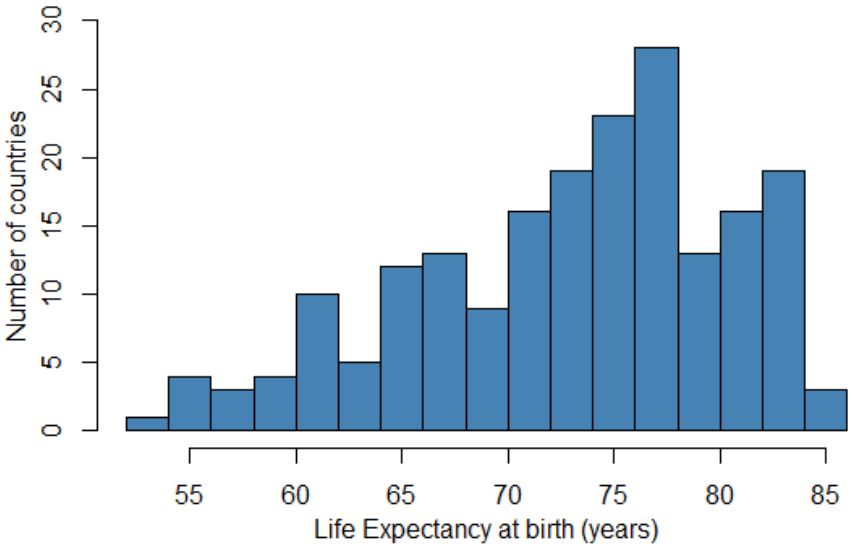
1. Preliminary analysis of the dataset

Overview of the data: The project uses the World Development Indicators (WDI) dataset. It contains data for 217 countries (rows) and contains measurements for 26 indicators (columns) such as GDP per capita, Percentage of people with access to electricity and other relevant indicators for the development level of each country.

Around 35% of the entries are missing. A breakdown of the missing values by column can be found in the appendix [2].

The predicted variable (Life Expectancy at birth) has a mean of 72.9 years and a standard deviation of 7.4 years. The country with the highest value is Hong Kong China, with a life expectancy of 85 years, while the one with the lowest value is the Central African Republic, with 53 years.

***Figure 1.1*: Histogram of Life Expectancy at birth**



As seen above, the distribution is left-skewed, which means that there are relatively few countries with low life expectancies, while the majority sits at values higher than 70 years. As it will be discussed in section 6, most of the low values come from countries in Africa.

2. Dealing with missing values

To build a linear model, the dataset that is used needs to have no missing entries. However, observations have on average 9.1 missing values with 0% of the rows being complete (or 0.9% if we exclude the empty column **c21**). For this reason, a complete case analysis cannot be done, so the alternative is to impute the missing data.

However, applying this method to columns with a high percentage of missing values should be avoided, since doing so might introduce bias into the system. Consequently, columns with more than 60% missing entries were removed, such as the following:

***Table 2.1*: Variables with more than 60% missing values**

Variable	c6	c7	c10	c21	c24
Percentage missing	83%	82%	88%	100%	89%

These 5 columns accounted for almost 50% of the missing values in the dataset.

Starting from the assumption that the data are missing at random, the method that was chosen to deal with this problem is **multiple imputation**, which is suitable for this dataset because it preserves variability, unlike mean imputation or other rudimentary alternatives. This is necessary due to the nature of the dataset.

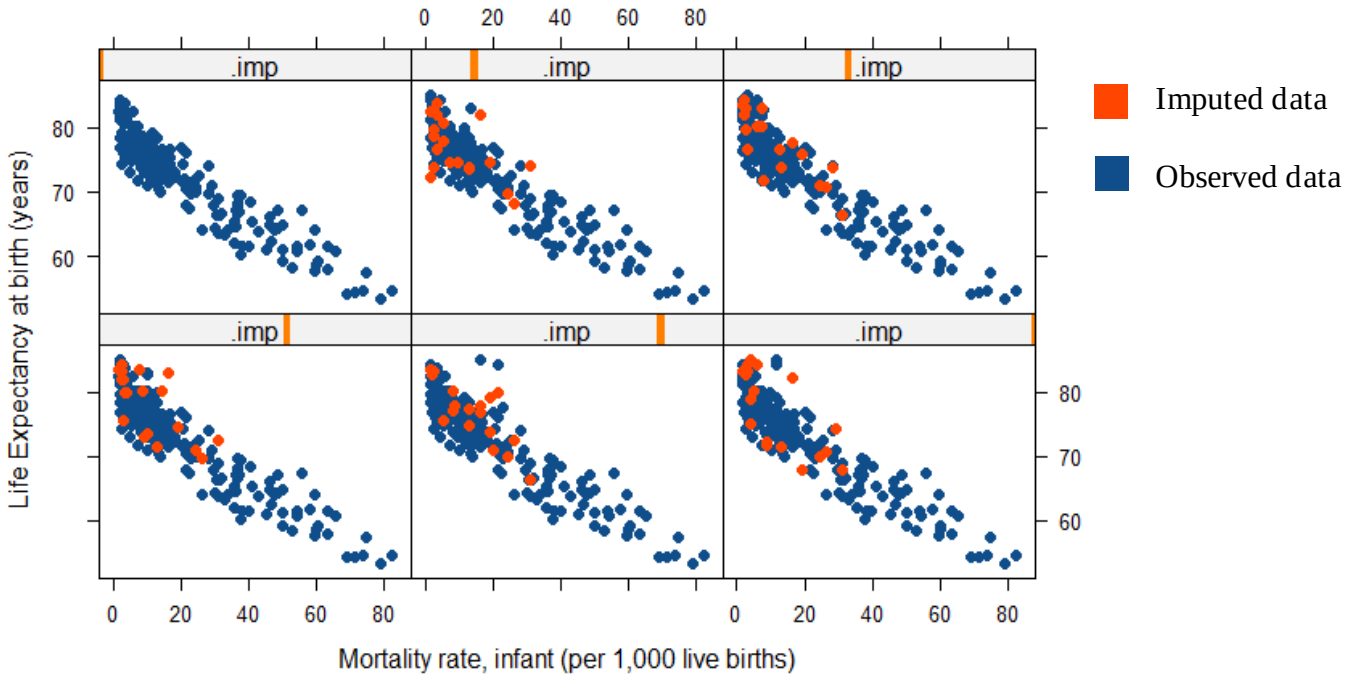
The method consists of constructing multiple copies of the dataset, each of which has its missing entries replaced by values from the observed distribution. The number of imputed datasets that was chosen for this analysis is 5.

Remark: Due to having diminishing returns, 5 copies is usually enough, which is why this number was chosen.

The values of **y** that were missing were imputes as well, since column **y** contains a relatively small number of missing values (≈8%), so even though the imputed data is not perfect, it was good enough to carry on with the analysis and not affect the final model.

To examine if the chosen method showed plausible results, the new data was plotted against the observed data for all five imputed datasets.

***Figure 2.2*: Imputed data plotted against the rest of the observations**



As shown in the graph above, the imputed data (shown in red) is plausible and follows the distribution of the observed data (shown in blue).

> **Result:** Examining similar graphs for multiple combinations of variables gave promising results, most of them showed a good approximation for the missing values. Therefore, there was enough confidence in the imputed datasets to be used for the next stage.

3. Dealing with collinearity

When assessing which factors should be included in the model, it is important to ensure there is little to no correlation between them. This idea is called *collinearity* and is known to cause problems for the model since it increases the variance of the coefficients.

To quantify the strength of the relationship between the predictors, two tools were used:

1. **The VIF** (*Variance Inflation Factor*) for each predictor variable. It is used to quantify the extent to which every predictor can be described by the rest of them.
2. **The correlation matrix** contains the pairwise correlation between the variables. It is used to identify pairs of highly correlated variables.

Remark: The analysis that is presented in this section was carried out on all of the 5 imputed datasets and the conclusion at the end is an aggregate of the results. However, for clarity, only the analysis for the 5th copy will be shown in full detail.

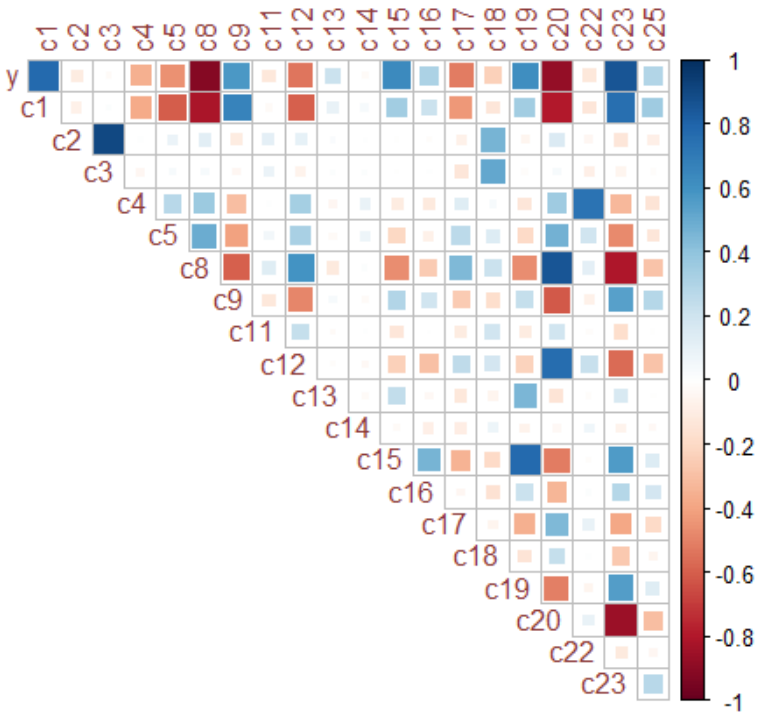
The consensus is that a VIF value greater than 5 is considered high ($R^2 > 0.8$), while a value over 10 is considered very high ($R^2 > 0.9$) and the corresponding variable should be removed from the dataset.

***Table 3.1*: The VIF of each variable for the 5th imputed dataset**

Variable	c1	c2	c3	c4	c5	c8	c9	c11	c12	c13
VIF	5.2	6.9	7.5	2.9	1.7	5.0	1.9	1.2	3.6	1.3
Variable	c14	c15	c16	c17	c18	c19	c20	c22	c23	c25
VIF	1.1	3.8	1.6	1.5	1.6	3.7	10.9	2.7	4.9	1.2

As illustrated above, there are a few variables with high and very high VIF. In particular, **c17** has a VIF of 10.9, which suggests that the information it contains is explained very well by the rest of the variables.

***Figure 3.2*: Visualization of the Correlation Matrix for the 5th imputed dataset**



As seen in Figure 3.2, multiple variables are correlated with each other. This result fits well with our intuition. For example, one would expect the **Adjusted net national income (c2)** to be correlated with the **Adjusted net national income per capita (c3)**, which is confirmed by the plot. Using both predictors should be avoided since one of them is most likely redundant.

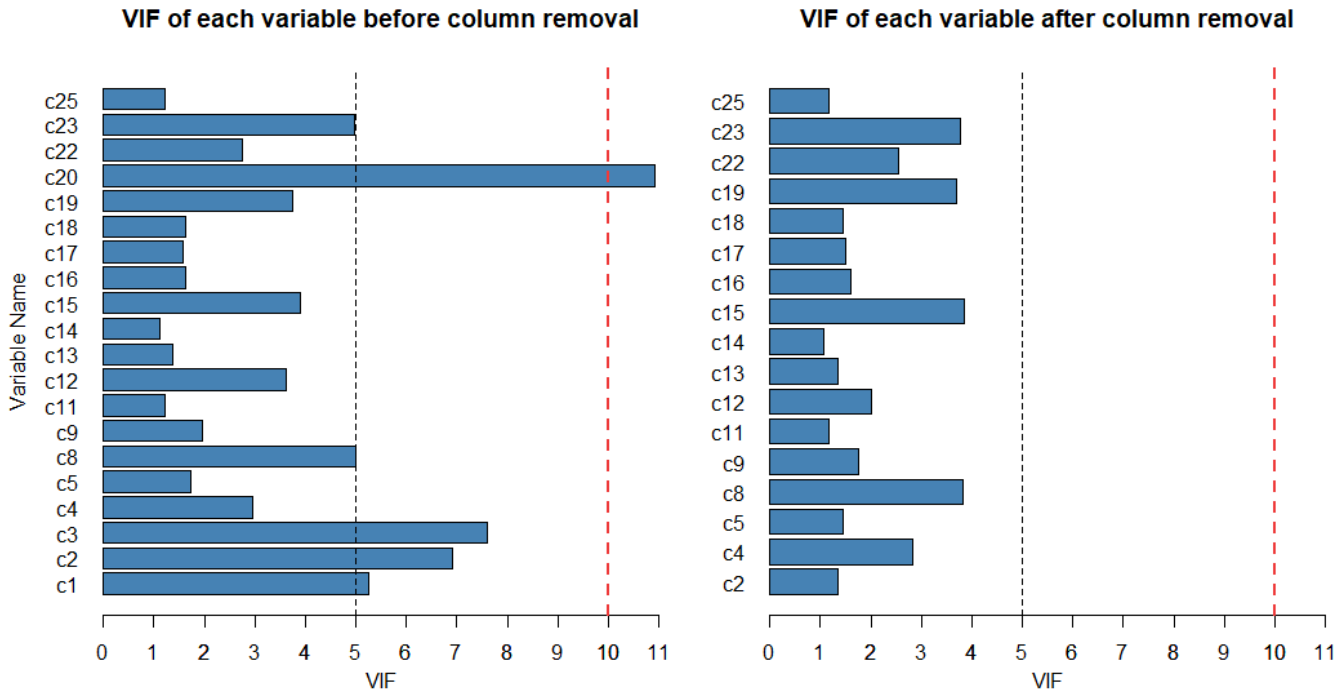
To reduce collinearity, variables were removed *one by one* based on the correlation matrix and the VIF table, using the following set of guidelines:

- Variables with very high VIF were given priority for removal (e.g. c17)
- If there were pairs of highly correlated variables, the member of each pair with a high VIF was considered for removal next (e.g. c2-c3**:** *0.91*, c20-c23: *-0.85*)
- Variables that show at least a moderate correlation with multiple others were removed (e.g. c1)

As much as possible, variables strongly correlated with **y** were kept in the model if they don't fall in one of the categories mentioned above, e.g. c8 (*-0.91 correlation with **y***) and c23 (*0.85 correlation with **y***).

Based on the set of guidelines shown above, **c20**, **c3**, **c1** were removed one by one until the VIF of the remaining columns was lower than 5.

***Figure 3.3*: The effect of removing c17, c3 and c1 on the overall VIF**



As seen in Figure 3.3, the removal of the three columns had a favourable effect on the overall VIF and was effective in dealing with collinearity.

Result: After carrying out a similar analysis on the other four imputed datasets, the majority of them displayed similar statistical characteristics and suggested the same variables to be removed. Therefore, **c17**, **c3** and **c1** were removed from the analysis, since keeping them causes collinearity.

4. Finding the best model

To find the best model to predict Life Expectancy, an iterative approach was employed **on each of the five imputed datasets**. This was done in three stages:

I. Selecting the variables

Starting from five initial models containing the intercept only (one for each dataset), a stepwise regression algorithm was applied to find five potential candidate models. The selection criterion was based on the AIC (Akaike information criterion) value, so at each iteration, the modified models with the lowest AIC values replaced the current ones.

Aiming to combine the five candidate models into one that was a good fit overall, the following table containing the frequency of each selected variable was devised:

***Table 4.1*: Frequency of each variable among the five candidate models**

Variable	c8	c23	c15	c22	c16	c17	c4	c12	c13	c19	c5	c11
Frequency	5	5	4	4	3	3	2	2	2	2	1	1

As illustrated above, variables **c8**, **c23**, **c15** and **c22** were selected in most cases, which suggests they are highly relevant for predicting **y**. Therefore, they were included in the unified model.

A further filtering of the variables was performed on the imputed datasets using the p-values, which suggested c22 is not significant at a 0.05 significance level, so it was removed from the set of predictors.

II. Evaluating the models

The model was fitted to each of the five datasets separately which resulted in an average R^2 of 0.89, while having significantly low p-values for all of the three variables. This means that the model is a good fit overall.

The model was also fitted on the 107 observations from the original dataset that contain no missing values for **c8**, **c15** and **c23**, including the predicted variable **y**, which gave an R^2 value of 0.9.

Remark: This last result is valuable since, as mentioned earlier, data is assumed to be missing at random, and a random subset of the observations should be representative of the entire dataset.

The normality of the residuals is one of the main assumptions of the model, so it was inspected in two ways:

- A numerical approach through a Shapiro–Wilk test: a hypothesis test used to determine if a variable follows the normal distribution.

  Using this method, the result was that the distribution of the residuals is not significantly different from the normal distribution for 4 out of the 5 models, as well as for the one based on the 107 complete observations at a 0.05 level.

- A visual approach through Q-Q plots

Q-Q plots were created for all fitted models, most of which followed the q-q line well, which is a good indication of the normality of the residuals (example in the appendix [4]).

III. Pooling (Combining the models)

To get the the predictors' coefficients for the unified model, the average of each coefficient across the 5 models was taken.

This resulted in the final model with the following coefficients:

***Table 4.3*: Coefficients of the pooled model**

Variable	Intercept	c8	c23	c15
Coefficient	74.1975189815	-0.2637189928	0.0492452611	0.0006708319

Variables c8, c15 and c23 refer to **Mortality rate, infant (per 1,000 live births)**, **Current health expenditure per capita** and the **Percentage of people using safely managed drinking water services** respectively.

Brief discussion on causality: This result comes as no surprise, since the higher the infant mortality rate is, the lower life expectancy at birth will be. Similarly, the development level of the health system is closely related to the conditions in which people are born, so naturally, the health expenditure per capita has a great impact on life expectancy. This ties in well with the overall focus of each country on the wellbeing of its population, which, among other things, means having clean water.

Result: The analysis suggests that a model consisting of **c8**, **c15** and **c23** is the best for predicting Life Expectancy at birth. It has the following form:

$$y = 74.1975189815 - 0.2637189928 \cdot c8 + 0.0006708319 \cdot c15 + 0.0492452611 \cdot c23.$$

To predict **y**, the values for the measurements of **c8**, **c15** and **c23** have to be multiplied by the coefficients and added together with the intercept. Variable **y** is correlated positively with c15 and c23 and negatively with c8.

The model is intuitive and, causally speaking, the chosen predictor variables seem plausible.

5. Predicting Life Expectancy

To estimate the life expectancy for the 19 countries where the value was missing, the model specified above was used on each of the five imputed datasets. This resulted in five different predicted values for each country which were then averaged out to get a final estimate.

Remark: Some of the countries whose life expectancies were predicted had multiple predictor variables missing in the original dataset. Therefore, the estimates might not be accurate for those cases.

***Table 5.1*: Predicted Life Expectancy (years) and missing variables for each country rounded to 1 digit**

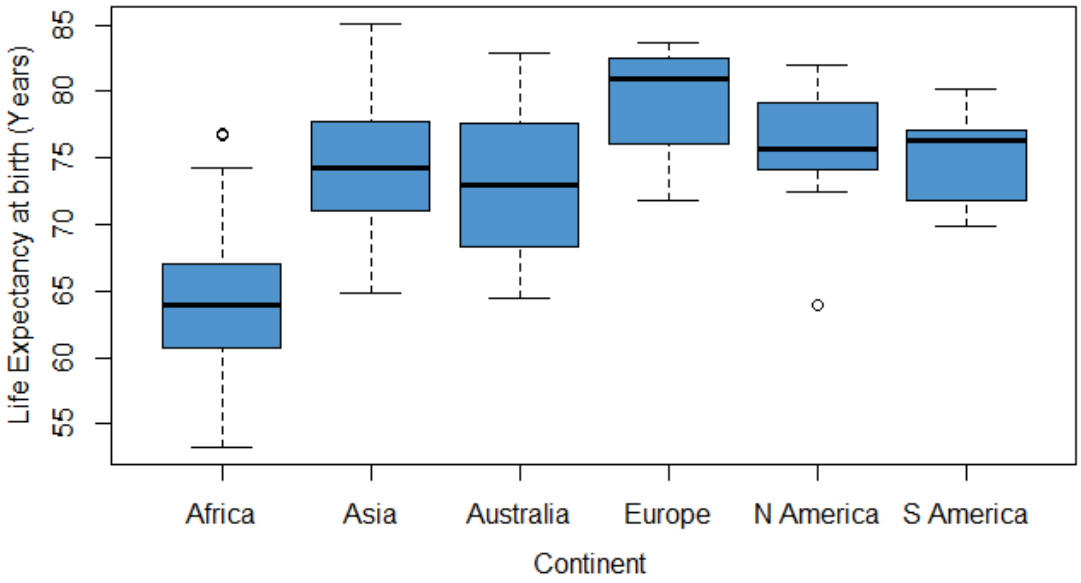
Country	Predicted	Missing	Country	Predicted	Missing
American Samoa	76.9	c8,c15	Monaco	80.4	none
Andorra	79.8	none	Nauru	71.1	c23
British Virgin Islands	74.9	all	Northern Mariana Islands	77.4	c8,c15
Cayman Islands	81.7	all	Palau	75.9	none
Curacao	77.1	all	San Marino	80.7	none
Dominica	70.0	c23	Sint Maarten (Dutch)	78.9	all
Gibraltar	79.5	c8, c15	St. Kitts and Nevis	75.8	c23
Greenland	79.7	c8, c15	Turks & Caicos Islands	75.9	all
Isle of Man	80.3	c8,c15	Tuvalu	73.5	c23
Marshall Islands	70.0	c23	-	-	-

Result: The estimates seem reasonable and there are no obvious outliers

6. Life expectancy across continents

Inspecting the original dataset, it might be helpful to understand if the continent in which a country is located says anything about life expactancy.

***Figure 6.1*: Boxplot of life expectancy at birth by continent (original dataset)**



Remark: Australia/Oceania was renamed Australia in the plot, due to the lack of horizontal space.

A visual inspection of Figure 6.1 indicates that there might be statistically significant differences between the mean life expectancy on different continents. In particular, Africa seems to have a much lower mean compared to the other five (exact values in the appendix [5]).

One – Way ANOVA

A One - Way ANOVA test was performed on the given data, which suggested that the means are different across continents at a 0.01 significance level.

However, this method is based on the assumptions that the residuals follow a normal distribution and that the variance is constant between the groups (continents). Even though the normality was evaluated using a Shapiro–Wilk test, which confirmed the assumption at a 0.05 significance level, a Bartlett's test showed that the variances are not homogenous, so a One – Way ANOVA model cannot be used.

Pairwise Comparisons: Dunnett's Modified Tukey-Kramer test

To further explore the differences, a DTK test was employed. The reason why this type of test was chosen instead of a more common option like a *Bonferroni post-hoc test* is because using it doesn't require having equal variance between the groups. This method provides confidence intervals for the pairwise differences between the means, which were used to decide if they were significantly different from 0, or in other words, if the means are different across continents.

> **Result**: The findings showed that at a 0.05 significance level, the mean life expectancies in Africa and Europe are significantly different from that of the other continents, as well as between themselves. In contrast, data suggests there is no difference between the means of Asia, Australia/Oceania, North America and South America.

Conclusion

The analysis indicates that life expectanty at birth is tightly related to the **infant mortality rate, current health expenditure per capita** and the **percentage of people using safely managed water services**.

As the model evaluation shows, these three measurements are effective at predicting life expectancy for each country and they can be used to estimate the values where the data are missing.

In general, countries with high infant mortality rates have lower life expectancies than those with lower infant mortality rates. Moreover, countries that spend more money on health and focus on providing the population with safely managed water services have on average higher life expectancies.

A discussion about the causality of these three factors concluded that there exists a natural explanation for their effect on life expectancy at birth.

The data also suggests that there exist statistically significant differences between the means of life expectancies across continents at a 0.05 significance level, Africa having the lowest one (64.1 years), and Europe having the highest value (79.2 years).

Appendix

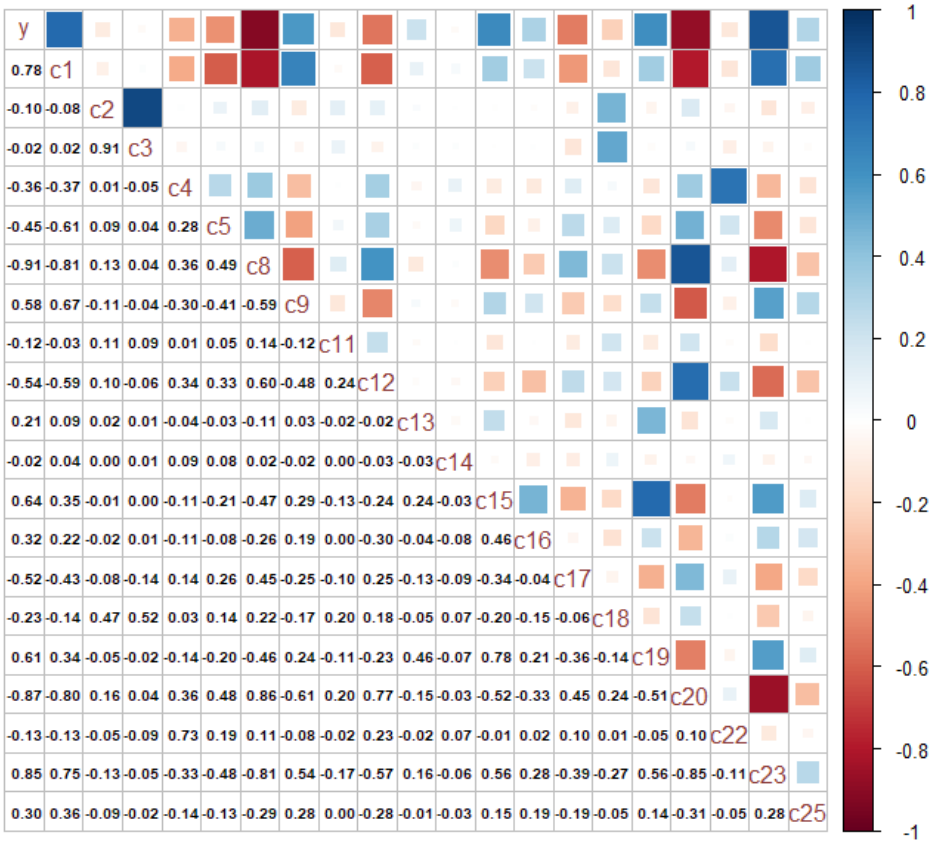
1.Names of columns

Name	Meaning	Name	Meaning
y	Life expectancy at birth, total (years)	c13	Population density (people per sq. km of land area)
c1	Access to electricity (\% of population)	c14	Population, total
c2	Adjusted net national income (annual \% growth)	c15	Current health expenditure per capita, PPP (\$)
c3	Adjusted net national income per capita (annual \% growth)	c16	Current health expenditure (\% of GDP)
c4	Children (ages 0-14) newly infected with HIV	c17	Unemployment, total (\% of total labor force)
c5	Children out of school, primary	c18	GDP growth (annual \%)
c6	Educational attainment, at least primary school, population 25+ years, total (\%)	c19	GDP per capita, PPP (\$)
c7	Educational attainment, at least Bachelor's, population 25+, total (\%)	c20	Birth rate, crude (per 1,000 people)
c8	Mortality rate, infant (per 1,000 live births)	c21	Renewable energy consumption (\% of total)
c9	Primary completion rate, total (\% of relevant age group)	c22	Adults (ages 15-49) newly infected with HIV
c10	Literacy rate, adult total (\% of people ages 15 and above)	c23	People using safely managed drinking water services (\%)
c11	Real interest rate (\%)	c24	Poverty headcount ratio at \$3.20 a day (2011 PPP) (\%)
c12	Population growth (annual \%)	c25	Compulsory education, duration (years)

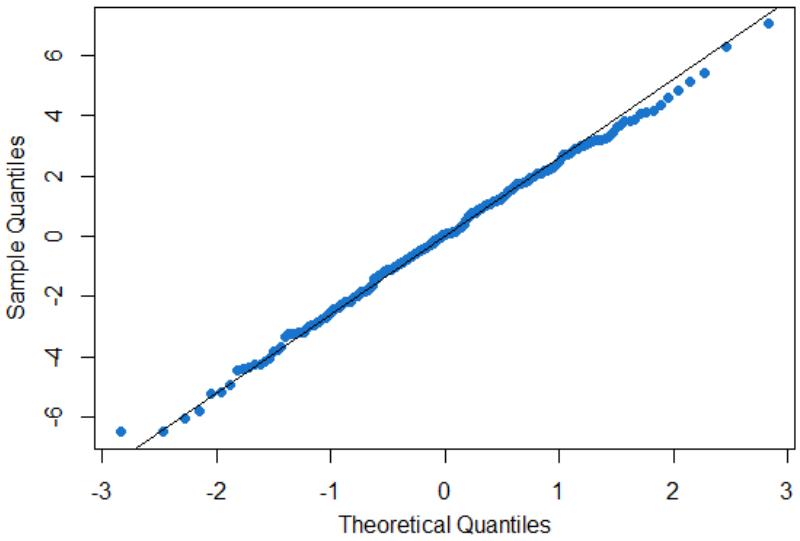
2. Breakdown of missing values by column (rounded to the closest integer)

Column	y	c1	c2	c3	c4	c5	c6
Missing	9% (19)	1% (1)	36% (79)	36% (79)	59% (127)	46% (99)	83% (181)
Column	c7	c8	c9	c10	c11	c12	c13
Missing	82% (179)	11% (24)	41% (89)	88% (192)	48% (104)	1% (1)	1% (1)
Column	c14	c15	c16	c17	c18	c19	c20
Missing	1% (1)	14% (31)	14% (31)	44% (96)	6% (14)	6% (12)	6% (13)
Column	c21	c22	c23	c24	c25	-	-
Missing	100% (217)	41% (88)	41% (89)	90% (195)	9% (19)	-	-

3. Full correlation matrix (with values)



4. Q-Q plot of the residuals for the third imputed dataset



5. Mean of life expectancy at birth (years) for each continent rounded to 1 digit

Continent	Africa	Asia	Australia	Europe	N. America	S. America
Mean	64.1	74.6	73.5	79.2	76.1	75.0

Prerequisites for the R code

```
if (!require("olsrr")) install.packages("olsrr")
if (!require("leaps")) install.packages("leaps")
if (!require("mice")) install.packages("mice")
if (!require("readr")) install.packages("readr")
if (!require("corrplot")) install.packages("corrplot")
if (!require("faraway")) install.packages("faraway")
if (!require("car")) install.packages("car")
if (!require("DTK")) install.packages("DTK")
library("readr")
library("olsrr")
library("leaps")
library("mice")
library("corrplot")
library("faraway")
library("car")
library("DTK")
```

```
#import data
Life_Expectancy_Data1 = read_csv("Life_Expectancy_Data1.csv")

#change names of the variables
names.equiv=rbind(names(Life_Expectancy_Data1), c("Country_Name",
"Country_Code", "Continent", "y","c1","c2","c3", "c4","c5","c6","c7",
"c8","c9","c10","c11", "c12","c13","c14","c15", "c16","c17","c18","c19", "c20",
"c21", "c22", "c23", "c24", "c25"))
names.equiv
names(Life_Expectancy_Data1)=c("Country_Name", "Country_Code", "Continent",
"y","c1","c2","c3", "c4","c5","c6","c7", "c8","c9","c10","c11",
"c12","c13","c14","c15", "c16","c17","c18","c19", "c20", "c21", "c22", "c23",
"c24", "c25")
```

R code for section 1. Preliminary analysis of the dataset

```
#check the data
head(Life_Expectancy_Data1)
dim(Life_Expectancy_Data1)
md.pattern(Life_Expectancy_Data1)

#check number of missing values per row
r_missing=matrix(0,nrow(Life_Expectancy_Data1),1)
for (i in 1:nrow(Life_Expectancy_Data1)){
     r_missing[i,1]=round(length(which(
                        is.na(Life_Expectancy_Data1[i,]))),2)
}
r_missing
mean(r_missing) #average missing values per row
sd(r_missing) #standard deviation of missing values per row
#number of complete observations
full_obs=0
for (i in 1:length(r_missing)){
  if (r_missing[i,1]==0){
    full_obs=full_obs+1
  }
}
full_obs
```

```
#breakdown of missing values by column
missing=matrix(0,3,ncol(Life_Expectancy_Data1))
missing[1,]=names(Life_Expectancy_Data1)
for (i in 1:ncol(Life_Expectancy_Data1)){
        missing[2,i]=round(100*length(which(is.na(Life_Expectancy_Data1[,i])))
                                                /nrow(Life_Expectancy_Data1),2)
        missing[3,i]=length(which(is.na(Life_Expectancy_Data1[,i])))
}
missing #contains count and percentage of missing values per column

#check means and standard deviation of y
mean(unlist(Life_Expectancy_Data1[!is.na(Life_Expectancy_Data1$y),4]))
sd(unlist(Life_Expectancy_Data1[!is.na(Life_Expectancy_Data1$y),4]))

#histogram of y
hist(Life_Expectancy_Data1$y, col="steelblue", breaks = 20, ylim=c(0,30),
                main="", xlab = "Life Expectancy at birth (years)", ylab="Number
                                                                of countries")
```

R code for section 2. Dealing with missing values

```
#deleting the columns where more than 60% values are missing
col_removed=c()
for (i in 1:ncol(Life_Expectancy_Data1)){
  if (as.numeric(missing[2,i])>=60){
    col_removed=cbind(col_removed, c(i))
  }
}
col_removed
Life_Expectancy_Data1=Life_Expectancy_Data1[,-c(col_removed)]

#imputation stage
imputations = mice(Life_Expectancy_Data1, method ="cart", seed = 7448)
imputations$imp #diagnosing issues e.g. negative numbers
#check quality of imputations
#same plot was created for multiple combinations of variables
xyplot(imputations, y ~ c8| .imp, col = c("dodgerblue4", "orangered1"),
                    pch = 20, cex = 1.3,ylab="Life Expectancy at birth (years)",
                          xlab="Mortality rate, infant (per 1,000 live births)")
```

R code for section 3. Dealing with collinearity

```
#select one of the imputed datasets
#same procedure was applied for all of them
reduced_df<-complete(imputations,5)[,-c(1,2,3)] #remove the character columns
that we don't want to check

#plot correlation matrix
corrplot(cor(reduced_df), method = 'square', type = 'upper', diag = FALSE,
tl.pos = "td", tl.cex = 1, tl.offset = 0.3,tl.col = "indianred4")
vif(reduced_df[,-1]) #check the VIF of each predictor

#Visualize the VIF
barplot(vif(reduced_df[,-1]), horiz = TRUE, las=1, col = "steelblue", main ="VIF
of each variable before column removal", xlab = "VIF", ylab = "Variable
Name",xlim = c(0,11))
axis(1, at = seq(1, 11, by = 1), las=1)
abline(v = 10, lwd = 2, lty = 2, col = "brown2")
abline(v = 5, lwd = 1, lty = 2)

#remove problematic columns one by one and check the VIF and correlation matrix
after every iteration
reduced_df=subset(reduced_df, select = -c(c20))
reduced_df=subset(reduced_df, select = -c(c3))
reduced_df=subset(reduced_df, select = -c(c1))
```

R code for section 4. Finding the best model

```
feature.selection1 = expression(null.model1 <- lm(y ~ 1),
                                model2 <- step(null.model1, scope    =
~c2+c4+c5+c8+c9+c11+c12+c13+c14
+c15+c16+c17+c18+c19+c22+c23+c25 ))
step.fit = with(imputations, feature.selection1) #stepwise regression
step.fit.models = lapply(step.fit$analyses, formula)
step.fit.features = lapply(step.fit.models, terms)
feature.frequency = unlist(lapply(step.fit.features, labels))
feature.frequency
sort(table(feature.frequency),decreasing=TRUE) #create frequency table

#check model for each dataset
summary(lm(data=complete(imputations,5), y ~ c8 + c23 + c15 + c22))
model.fit <- with(imputations, lm(y ~ c8 + c23 + c15))
summary(model.fit)
pooled.model<-pool(model.fit)
summary(pooled.model) #get coefficients for the pooled model


#test for normality of the residuals for each fitted model (change n to check
the n-th model)
n=5
qqnorm(residuals(lm(formula = y ~ c8 + c23 + c15, data = complete(imputations,
                                  n))), pch = 20, cex=1.4, col="dodgerblue3",
                          main="Original dataset",ylab="", xlab="", xaxt="n")
qqline(residuals(lm(formula = y ~ c8 + c23 + c15, data = complete(imputations,
                                                          n))), lt = 1)

shapiro.test(residuals(lm(formula = y ~ c8 + c23 + c15,
                                              data = Life_Expectancy_Data1)))
```

R code for section 5. Predicting life expectancy

```
y_pred=rep(0, nrow(Life_Expectancy_Data1))
for (i in 1:5){
  model=lm(data=complete(imputations,i), y ~ c8 + c15 +c23)
  model$coefficients=c(74.1975189815, -0.2637189928, 0.0006708319, 0.0492452611)
  y_pred=y_pred + predict(model, complete(imputations,i))
}
y_pred=y_pred/5

#check dataframe with the predicted values and predictors
pred_df=cbind(Life_Expectancy_Data1$Country_Name, Life_Expectancy_Data1$y,
                  y_pred, Life_Expectancy_Data1$c8, Life_Expectancy_Data1$c15,
                                                    Life_Expectancy_Data1$c23)
```

R code for section 6. Life expectancy across continents

```
#create data frame with the necessary columns, omitting NA values
cont=data.frame(Country_Name=Life_Expectancy_Data1$Country_Name,
         Continent=Life_Expectancy_Data1$Continent, y=Life_Expectancy_Data1$y)
cont=na.omit(cont)
tapply(cont$y,cont$Continent,mean) #check means of each continent

#conduct a one-way ANOVA
one_way=aov(cont$y ~ cont$Continent)
summary(one_way)

#check normality of residuals
qqnorm(one_way$residuals, pch=19)
shapiro.test(one_way$residuals)
hist(one_way$residuals)

#check variance homogeneity
bartlett.test(cont$y ~ cont$Continent)

#perform DTK test
DTK.test(cont$y, cont$Continent, a = 0.05)
```

References

DTK test details: https://www.rdocumentation.org/packages/DTK/versions/3.5/topics/DTK.test

Bonferroni test assumes equal variances: https://www.ijntse.com/upload/1447070311130.pdf

Contributions

Group 50 was made up of Rareș Lițescu, [redacted] and [redacted].

To make it less cluttered, only the first names were used in the table below.

Task	Done by	Contributed
Introduction	Rareș, [redacted], [redacted]	-
Section 1	Rareș, [redacted]	-
Section 2	[redacted]	Rareș
Section 3	Rareș	-
Section 4	[redacted]	Rareș
Section 5	Rareș	-
Section 6	Rareș	-
Conclusion	Rareș, [redacted], [redacted]	-
Formatting	Rareș	-
PowerPoint	Rareș, [redacted]	

Thank you!